北京航空航天大学计算机学院

本科生毕业设计（论文）开题报告

论文题目：三维场景人体与对象交互的接触点标注工具

学生姓名： 李奕君

学生学号： 15061199

专 业： 计算机科学与技术

指导教师： 周彬

学院(系)： 计算机学院

北京航空航天大学计算机学院

**2018年 10月 30日**

**本科生毕业设计(论文)开题报告**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学生姓名 | 李奕君 | 学号 | 15061199 | 专业班级 | 150617 |
| 指导教师 | 周彬 | 职称 | 讲师 | 单 位 | 北京航空航天大学计算机学院 |
| 题目类型 | □理论研究 应用研究 □综述 □其它  如选“其它”，请自定义题目类型 | | | | |
| 题目来源 | 校内 □校外 | | | | |
| 毕业设计（论文）题目 | 三维场景人体与对象交互的接触点标注工具 | | | | |
| **开题报告（阐述课题的目的、意义、研究现状、研究内容、研究方案、进度安排、预期结果、参考文献等）**   1. **研究背景与意义**   本毕业设计项目研究工作来源于国家自然科学基金重点项目“感知素材可交互内容编辑与生成理论及方法”。  由于图像采集技术和多媒体技术进一步发展，每时每刻都涌现出大量的图像素材，并且在互联网及云技术的推动下得以被广泛下载和传播。如何方便快捷地从这些大量的数据中获得对人类有用的信息是一个有待研究的问题。当前计算机视觉领域理解图像内容的方式是研究和获取图像中的语义和属性等信息。如今，在计算机软硬件性能不断提升的背景下，简单地观赏和理解图像中的内容已经无法使人们得到满足，人们更希望能够借助虚拟现实技术将图像的内容变得更加生动形象甚至能够与人交互。  根据单幅图像找到人体和对象接触点能够帮助我们分析图像中人和对象的交互行为和接触位置，从而更好地还原三维交互场景，帮助计算机理解如何与一个物体进行交互。  **二、国内外研究现状**  国内外研究现状内容主要包括：图像中交互关系解析方法，图像中人体二维姿势估计方法，图像中人体三维姿势估计方法和基于图像的三维对象建模等相关方法。  2015年密歇根大学构建了人与常见对象交互关系标准知识库，对于知识库中每一张图像标注了人与对象的交互关系。该知识库包含520个动词-对象关系对，116个交互动词，80个对象类别以及“没有交互”类别和600个人体与对象交互关系类别，总共标注了47774张图像。  2016年伯克利大学提出一种使用反馈信息构建的自动处理复杂的、结构化输出的空间预测器。通过该方法输入一张图像能够得到图像中人物的一些二维关节点。  2016年德国波恩大学利用大量训练数据从单张图像中估计人体三维姿势。该方法用到了两种不同类型的数据库，一类含有二维姿势标注的图像数据；另一类是精确的三维运动捕获数据。该方法将二维姿势估计与三维姿势检索相结合从而估计单幅图像的人体三维姿势。  2009年美国布朗大学提出一种利用单幅图像估计其中人体三维姿势的方法，该方法需要用户标注出图中人体二维姿势关节点位置和人的身高信息，然后利用缩放正交投影假设估计出人体的三维姿势，结合SCAPE模型和身高参数获得人体的三维模型。  2015年斯坦福大学构建了一个能够同时包含二维图像与三维模型的联合空间，该空间中实体节点间的距离表示了图像中对象与三维模型的相似度。使用同一个三维模型在不同视角和不同背景下渲染大量图像，再用卷积神经网络(CNN)对图像信息提纯得到一个特征向量，该特征向量即为联合空间中的一个实体节点。对于输入的一张图像，通过神经网络获取其特征并映射到共存空间，与映射节点的距离最近的节点所代表的三维模型即为最相似的三维模型。  **三、研究目标**  本次毕业设计的目标在于能够自主开发一款三维场景人体与对象交互的接触点标注工具。对于这个工具而言，输入是一张人与物体的交互图片以及该图片对应的三维场景原始文件，其中文件里包括多个人或物体的三维模型。导入三维模型后，该工具能够通过图片自动调整人和物体三维模型的初始位姿，使用者可以通过键盘或鼠标进一步调整模型的位姿和大小。三维模型按照照片所示放置好以后，工具能够通过计算给出初步的人与物体接触点的估计，在此基础上使用者可以通过该工具提供的方式进一步修正三维接触点的位置、大小以及数量，从而达到三维场景中人和物体接触点标注任务的实现。  通过此次毕业设计研究三维场景人体与对象接触点的标注能够帮助我们分析图像中人和对象的交互行为和接触位置，从而更好地还原三维交互场景，进一步帮助计算机理解现实世界中一个真实物体该如何使用。  **四、主要研究内容**  在做毕设过程中，本人将分为三步实现三维场景人与对象的接触点标注工具：首先实现使用者能够通过导入三维场景并通过本人设计的一种有效算法让标注工具通过自主计算的方式求解三维场景中的人与物体的接触点信息。其次实现使用者能够通过手工标注的方式，在给定计算好的标注点的基础上进一步调整接触点的大小、位置及数量等信息。最后实现使用者能够从标注工具相关联的三维模型数据集中检索图像中的物体并自动生成的人物模型的功能，再将模型导入图像和对应的三维场景，在调整好位姿和大小后满足使用者标注三维接触点的需求。   1. 支持自动求解三维接触点   本人打算研究一种有效的算法能够让标注工具来较为准确地求解三维场景中人与物体三维模型的三维接触点。该算法可以从三维模型的结构、组成方式和三角面片分布入手。该工具支持使用者手工导入单幅图像以及对应的三维场景模型，使用者可以让标注工具通过算法给出初步的接触点分布。   1. 支持手工标注三维接触点   在完成支持使用者通过自动求解方式标注三维模型的三维接触点以后，使用者可以在给出的初步接触点分布下通过工具手工添加或删除一个或多个三维接触点，调整接触点的大小和位置，从而完成该三维场景的人和物体三维接触点标注工作。   1. 支持检索物体和生成人物模型来构成三维场景并计算接触点   使用者用该工具打开一张人与物体交互图片，在此基础上能够在三维模型数据集中手动检索图片中物体对应的三维模型并导入工具中的三维场景中，然后调整生成好的三维人物模型的姿势，最后按照图片上的三维模型投影调整好人和物体模型的位置和大小实现三维接触点的标注任务。采用的物体三维模型数据集是ObjectNet3D。   1. 撰写毕业设计论文   根据实现的标注工具和收集的资料撰写毕业设计论文，描述三维场景人与物体的接触点标注工具的具体流程，总结该工具开发时遇到的技术难点及解决措施以及使用过程中遇到的问题和相应解决方法，并附上致谢和参考文献等其它相应内容。  **五、拟采取的研究方案及可行性分析**  1、研究方案  在开发本标注工具的过程中，本人将制定明确的研究规划表，并查阅相关文献，总结开发经验，对相关和相似的工作进行考究，通过不断测试检查标注工具合理性和鲁棒性，对其进一步完善，并使其更加人性化。在此基础上参考指导老师意见，并进一步优化标注工具， 使其功能更加完备。  (1)制定研究规划表：将实现三维场景人与对象的接触点标注工具的任务分成若干阶段的子任务，并对划分好的子任务规定完成时间段，每完成一个子任务后通过检查和测试的方式确保该任务达标后再进行下一阶段的任务开展。  (2)查阅相关资料：积极查阅相关资料，将别人做好的对该标注工具实现有利的工作进行总结归纳，吸收其中的精妙思想，用新颖高效地方式来实现标注工具，加快标注工具的开发进度，提高开发效率。  (3)结合指导老师建议修正：在实现标注工具的过程中，积极与指导老师沟通，对出现的问题及时解决，对思考欠妥的地方进行更深入的考察，更好地提高工具的实用性，使标注工具更加健全和完善，使用更加方便。  下面是本文的毕设大致流程图：  2、可行性分析  本项目具有较大的可行性，具体原因如下：   1. 项目所需资源来源明确可靠，而且易于收集。 2. 项目负责人经验丰富，编程能力强，能够熟练使用多种相关软件技术对项目进行开发。 3. 指导老师热心负责，能够及时提供改进建议。 4. 毕设实现流程明晰，可行性高。   **六、预期目标和成果**  完成一个三维场景人体与对象交互的接触点标注工具，使用者可以输入单幅图像和对应的三维场景，并利用该工具对三维场景中的人和对象模型进行调整，通过自动求解和手动标注的方式完成三维接触点标注任务。该工具还能支持检索物品三维模型和生成人物三维模型的需求。  **七、论文工作计划**   |  |  |  | | --- | --- | --- | | 开始时间 | 结束时间 | 完成工作 | | 2018年11月1日 | 2018年11月10日 | 制定研究规划表 | | 2018年11月11日 | 2018年11月20日 | 查阅收集相关资料 | | 2018年11月21日 | 2018年12月21日 | 实现自动求解三维接触点功能 | | 2018年12月22日 | 2019年1月15日 | 实现手工标注三维接触点功能 | | 2019年1月16日 | 2019年3月31日 | 实现物体模型检索和人物模型生成功能 | | 2019年4月1日 | 2019年4月30日 | 对标注工具进行测试和完善 | | 2019年5月1日 | 2019年6月 | 准备毕设结题论文 |   **八、主要参考文献**  [1] Yu Xiang, Wonhui Kim, Wei Chen, Jingwei Ji, Christopher Choy, Hao Su, Roozbeh Mottaghi, Leonidas Guibas and Silvio Savarese. ObjectNet3D: A Large Scale Database for 3D Object Recognition. 2016  [2] Marszalek M, Schmid C. Semantic Hierarchies for Visual Object Recognition. 2007  [3] Fellbaum C, Miller G. WordNet:An Electronic Lexical Database. 1998  [4] Chao Y W, Wang Z, He Y, et al. HICO: A Benchmark for Recognizing Human-Object Interactions in Images. 2015  [5] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts. 2011  [6] Carreira J, Agrawal P, Fragkiadaki K, et al. Human Pose Estimation with Iterative Error Feedback. 2016  [7] Camillo Jose Taylor. Reconstruction of Articulated Objects from Point Correspondences in a Single Image.2000  [8] Guan P, Weiss A, Bãlan A O, et al. Estimating Human Shape and Pose from a Single Image. 2009  [9] Li Y, Su H, Qi C R, et al. Joint embeddings of shapes and images via CNN image purification. 2015  [10] Ionescu C, Papava D, Olaru V, et al. Human3.6M[DB/OL]. 2014  [11] Girshick R. Fast r-cnn. 2015  [12] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014  [13] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database. 2009  [14] Valmadre J, Lucey S. Deterministic 3D Human Pose Estimation Using Rigid Structure. 2010  [15] CMU. Graphics Lab Motion Capture Database[DB/OL]. 2007  [16] Chao Y W, Wang Z, He Y, et al. HICO: A Benchmark for Recognizing Human-Object Interactions in Images. 2015 | | | | | |